High-speed 3D sensing via hybrid-mode imaging and guided upsampling

High-speed 3D sensing via hybrid-mode imaging and guided upsampling

ISTVAN GYONGY[1,*], SAM W. HUTCHINGS[1], ABDERRAHIM HALIMI[2], MAX TYLER[2], SUSAN CHAN[2], FENG ZHU[2], STEPHEN MCLAUGHLIN[2], ROBERT K. HENDERSON[1], AND JONATHAN LEACH[2]

[1] School of Engineering, Institute for Integrated Micro and Nano Systems, The University of Edinburgh, Edinburgh, EH9 3FF, UK
[2] School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, UK
[*] istvan.gyongy@ed.ac.uk

Compiled July 27, 2020

Imaging systems with temporal resolution play a vital role in a diverse range of scientific, industrial, and consumer applications, e.g. fluorescent lifetime imaging in microscopy and time-of-flight (ToF) depth sensing in autonomous vehicles. In recent years, single-photon avalanche diode (SPAD) arrays with picosecond timing capabilities have emerged as a key technology driving these systems forward. Here we report a high-speed 3D imaging system enabled by a state-of-the-art SPAD sensor used in a hybrid imaging mode that can perform multi-event histogramming. The hybrid imaging modality alternates between photon counting and timing frames at rates exceeding 1000 frames per second, enabling guided upscaling of depth data from a native resolution of 64×32 to 256×128. The combination of hardware and processing allows us to demonstrate high-speed time-of-flight 3D imaging in outdoor conditions and with low latency. The results indicate potential in a range of applications where real-time, high throughput data is necessary. One such example is improving the accuracy and speed of situational awareness in autonomous systems and robotics.

http://dx.doi.org/10.1364/optica.XX.XXXXXX

1. INTRODUCTION

Three-dimensional depth sensing is used in a growing range of applications, including autonomous vehicles [1], industrial machine vision [2], gesture recognition in computer interfaces [3], and augmented and virtual reality [4]. Amongst a number of approaches to capture depth, time-of-flight (ToF) [5], which illuminates the scene with a modulated or pulsed light source, and measures the return time of the back-scattered light, is emerging as an appealing choice in many applications. Advantages include greater than centimetre depth resolution over distances ranging from a few meters to several kilometers. In contrast to alternative techniques such as stereoscopy [6] and structure-from-motion, there is low computational overhead, and no reliance on scenes being textured. Furthermore, ToF uses simple point, blade or flood-type illumination, as opposed to the projection patterns that structured light-type approaches rely on [7].

Whilst frame rates of 10-60 frames per second (FPS) are typical for ToF, an order of magnitude faster acquisition rates, coupled with minimal latency would be beneficial in several applications. In autonomous cars, for example, fast 3D mapping of the environment would help ensure the timely detection of obstacles. For city driving, video rate acquisition equates to several meters of travel for every few frames of 3D data, which may mean the difference between a collision being avoided or not. Similarly, augmented reality requires fast capture of the user's environment, so that it can be interpreted by computer vision, and digitally enhanced, in real time for a seamless experience. In a broader context, ToF at > 1 kFPS would access the realm of scientific imaging, and enable the recording of transient, high-speed phenomena [8], such as in fluid dynamics, not possible with current ToF technology.

Achieving high frame rates requires high photon efficiency

throughout the pipeline of converting incident photons into timing information as presented at the outputs of the sensor. Furthermore, the parallelised acquisition of 2D array format sensors, coupled with flood-type illumination [9], offers higher potential frame rates than systems based on a single-point sensor and beam steering, where the scanning rate can be a limiting factor. From the perspective of photon-efficiency, SPADs have inherent advantages, thanks to an ability to time individual photons with picosecond timing resolution, and shot-noise limited operation. However in SPAD image sensors providing time-correlated single-photon counting (TCSPC), both the fill-factor, and the overall photon throughput has been relatively low, compared to the maximal rate of > 100 M events/s that a single SPAD can generate [10]. This is due to the use of photon timers, or time-to-digital converters (TDC), which register only the first detected photon in every frame [11]. Whilst computational imaging approaches [12–14] have been proposed to estimate depth from sparse photon events, current approaches tend to be computationally intensive, and the "filling in" of gaps in data may not be acceptable in safety critical applications. A further disadvantage of "first photon" TDCs is a susceptibility to distortion in the resulting timing histograms under high ambient levels, corrupting depth estimates. A number of strategies have been proposed to reduce this distortion [15–17], the offsetting of the photon measurement window with respect to the laser cycle having been shown as the most effective approach [18].

Previously reported high-speed ToF results include underwater depth imaging [19] with a 192 × 128 SPAD at binary (first-photon) frame rates approaching 1 kFPS (the resulting depth frames showing relatively sparse depth information due to low photon returns). Another study [20] presents indoor depth results from a 32 × 32 InGaAs SPAD running at a binary frame rate of 50 kFPS. Frames are accumulated in groups of ≥25 and Kalman filtering applied to obtain depth maps, the example timing histograms provided showing evidence of pile-up effects. The same SPAD has been used to demonstrate a computationally efficient approach for reconstructing 3D scenes from single-photon data in real-time at video rates [21]. A frame rate of 200 FPS has been shown for a 64 × 32 SPAD with an indirect ToF architecture [22]. It is also important to mention compressive sensing ToF systems [23], that have the potential of generating depth maps with high frame rates, by reducing the number of measurements. However, at present, the reconstruction of frames can be computationally demanding.

In this work, we use a state-of-the-art SPAD array sensor [24] for high-speed 3D sensing. The sensor has a 3D-stacked structure with separate detector and photon processing tiers, that enables high-fill factor of 50% and an increased processing capability within the array. The array has a full resolution of 256 × 256 pixels, and this is made up of 64 × 64 macropixels, each containing a small array of 4 × 4 SPADs. The sensor can operate in multiple modes, two of which are relevant to this work: first, intensity or photon counting mode at a resolution of 256 × 256; and second, multi-event TCSPC histogram mode at a resolution of 64 × 64. To maximise the potential frame rate of the sensor, we halve the number of rows read out, thus doubling the frame rate.

In the intensity mode each pixel provides a 14-bit photon count, thus, in principle the photon counting capacity of the sensor is $256 \times 128 \times (2^{14} - 1) \approx 0.5$ giga photons in a single frame. In the histogram mode, events in each 4 × 4 macropixel are combined to provide a single histogram, hence the reduced resolution in this case. Each histogram contains 16 bins, and each bin has a minimum temporal resolution of 500 ps and a photon counting capacity of 14-bits. The temporal bin width of the sensor can be increased arbitrarily. When operating in histogram mode, the photon counting capacity of the entire sensor is $64 \times 32 \times 16 \times (2^{14} - 1) \approx 0.5$ giga photons in a single frame. The consequence of this is that the sensor is able to operate in high photon flux environments without getting saturated.

For this work, we have developed the firmware of the sensor with regards to [24] such that it can operate in a hybrid imaging mode at high speeds. In the hybrid imaging mode, high-resolution intensity images and low-resolution time-of-flight histograms can be captured in an interleaved fashion. The advantage of the hybrid imaging mode is that we have a high resolution intensity image with which to guide the upsampling of the lower resolution depth information, resulting in a four-fold improvement in the spatial resolution of the depth data.

The sensor operates such that alternating frames at ≈ 500 FPS in intensity and histogram mode are captured, providing an overall frame rate of ≈ 1 kFPS. The upper estimate of the maximum photon throughput of the sensor is then ≈ 500 giga photons per second. Table 1 compares the maximum photon counting in 1 ms of the sensor to other state-of-the-art devices. We see that the sensor used in this work has a maximum photon counting capacity of ≈ 500 mega photons in 1 ms. This is a three orders of magnitude improvement in total photon count, thus enabling operation in high photon flux environments.

The work presented here demonstrates high-speed 3D imaging in ambient light conditions. This is enabled by the unique combination of the factors mentioned above: first, the state-of-the-art SPAD array that can operate in a high photon flux environment; second, firmware that enables alternating modes of imaging at high rates; and third, the guided upsampling algorithm that upscales the native resolution of the depth data.

Figure 1 illustrates the advantages of multi-event histogramming over conventional first photon timing: photon-rich histograms are generated in-pixel, which dramatically increases the acquisition rate of photons. Furthermore, pile-up distortion is minimised, as it requires multiple photon detections within the time interval of a bin, rather than within the entire histogram time period, and when it does occur, its effect is independent of bin position. We also note that as the SPADs are continually active, rather being turned on at the start of the timing period, detector pile-up due to the SPAD dead-time and macro-pixel combination tree [25] does not distort towards early time bins either.

2. EXPERIMENT

The experimental setup is illustrated in Figure 2, and has the SPAD camera trigerring a pulsed, fibre-coupled laser source (Picoquant LDH-Series 670 nm laser diode, 60MHz repetition rate), whose light is spread over the fast-changing scene to be captured using a 3.3 mm, NA = 0.47 aspheric lens (Thorlabs N414TM-A). Imaging is through a 3.5 mm/f1.4 objective (Thorlabs MVL4WA, giving a 25 degrees diagonal field-of-view), resulting in matching imaging and illumination cones. Adjustable ambient illumination is provided by a high-intensity LED array. The 40 mW average optical power from the laser is sufficient for the setup to achieve sub-cm depth precision for targets at a close distance range (2 m) whilst maintaining high frame rates in the kFPS range. Global shutter is used, so that the camera frame rate is given by $1/(T_{exp} + T_{read})$ where T_{exp} is the exposure time and $T_{read} = 655\ \mu$s is the frame readout time. The total power

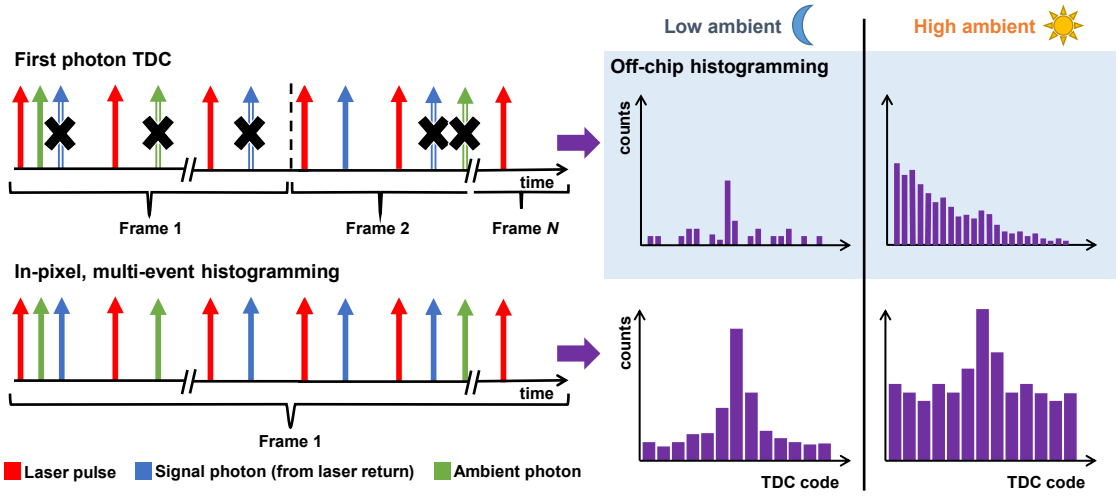

Fig. 1. Photon registration for a given pixel in a conventional direct ToF sensor (top plots) versus an in-pixel, multi-event histogramming sensor (bottom plots). A conventional sensor registers only the first photon detected in the frame time, which could be an ambient photon. Thus, multiple frames are required to build up a histogram of photon arrival times from which depth can be reliably estimated. Furthermore, the histograms, which are accumulated off-chip, become distorted in high ambient conditions, due to the dominance of early photons. In contrast, the present multi-event histogramming sensor is able to register multiple photons per frame, even within the same laser cycle, if falling into different bins. Hence the sensor can generate photon-rich histograms, in-pixel, for each frame. Orders of magnitude higher number of photons can be collected within the same acquisition time this way, and pile-up in high ambient is minimised.

Table 1. Comparison of direct ToF sensors used in high-speed imaging in terms of array pixel count A, number of bins n, bin size δ, frame rate f_{max}, number of photons N_{max} acquired in 1 ms

	[19]	[20]	this work
A	192 × 128	32 × 32	64 × 32
n	4096	8000	16
δ	33-120 ps	250 ps-1.25 ns	$\geq$ 500ps
f_{max}	binary @ 10 kFPS	binary @ 50 kFPS	14-bit histogram @ 1 kFPS
N_{max}	**246 k**	**51.2 k**	**500 M***

*theoretical; 197M measured outdoors with 400 μs exposure

consumption of the sensor is <100 mW.

Figure 2 also shows a sample depth frame from the SPAD when capturing a high-speed (1000 RPM) fan. The figure also gives the histogram corresponding to a macropixel, showing time resolved photon returns from the fan blade. The bin width in this case is around 700 ps. The histogram can be approximated as a sampled Gaussian function with a vertical offset, each bin being subject to Poisson noise. Depth may be obtained by estimating the time position of the peak using iterative curve fitting [26], but a simple approximate maximum likelihood method leads to similar performance. The latter reduces to a localized centre-of-mass method using signal counts, obtained after subtraction of background counts from the histograms [27]. A scenario where centre-of-mass gives sub-optimal results is when there are multiple overlapping peaks in the histogram.

To highlight the considerable photon throughput of the system Figure 3a shows an example depth frame, obtained in high ambient conditions, of a person juggling outdoors. The sequence was captured under the midday sun on a clear late-April day in Edinburgh, Scotland, leading to considerable solar radiation at the laser wavelength (670 nm). Despite the high ambient level, the content of the frame, i.e. the torso, arms, ball, is clearly recognisable. The figure also plots the histogram for a macropixel registering photons from the surface of the ball, indicating an ambient level of around 0.9 background photons per laser cycle. At such level of background photons, conventional first-photon TDCs suffer from considerable photon pile-up effects [15], making it difficult to detect the laser return and hence capture an accurate depth map, as illustrated using synthesised data in Figure 3b. We note that the multi-event TDC used here gives a histogram free from obvious distortions, as evidenced by the flat baseline, and a visible signal peak, despite the short 300 μs exposure time.

Depth frames captured using the camera are limited, in their native form, to the macropixel resolution of 64 × 32. However they may be upscaled, with relatively low computational needs, to the detector resolution of 256 × 128, by acquiring photon counting data, at this resolution, in alternate frames. The scheme, illustrated in Figure 4 has the following steps. The number of depth frames is upconverted to the overall frame rate of the camera to produce depth frames that are aligned with the intensity images. The newly generated depth frames are then upscaled according to the corresponding intensity data, and 3D images are generated from the resulting depth frames, with intensity overlaid. The overall processing time for a Matlab implementation running on a PC with Intel® Core™ i7-4790 CPU at 3.60 GHz and 32 GB RAM is currently in the 50 ms region. However, as significant portions of the algorithm operate on individual or groups of pixels, it is anticipated that with parallelisation, and potential hardware acceleration, the computational time can be

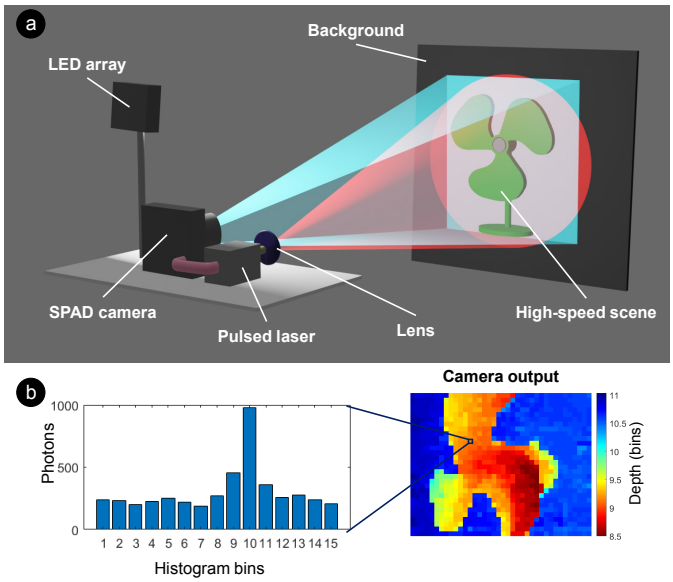

Fig. 2. (a) Setup for high-speed 3D data capture using the SPAD camera. A pulsed laser, triggered by the SPAD, illuminates the scene via an optical fibre followed by lens to divergence the beam. In addition, a high intensity, continuous wave, white LED source is used to provide ambient illumination indoors. (b) Example depth frame from SPAD, cropped to 32×32 pixels, together with the underlying time-correlated single-photon counting (TCSPC) histogram for a selected macropixel. Depth is obtained by peak extraction, followed by centre-of-mass calculation, on the histograms. The exposure time was 500 μs.

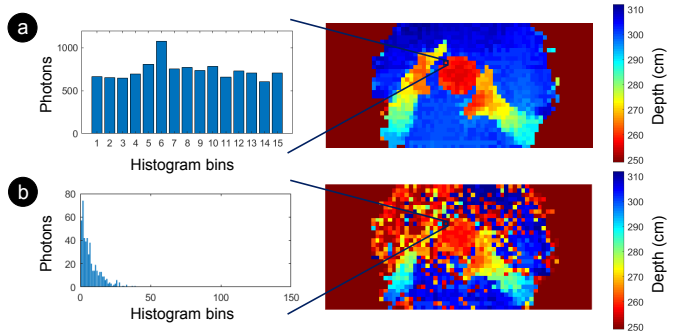

Fig. 3. a) Depth frame from outdoor juggling sequence, together with the underlying TCSPC histogram for a selected macropixel. The histogram shows a considerable background level without obvious photon pile-up effects. The background level corresponds to around $\approx$ 0.9 photons per laser cycle. b) A synthesised version of the depth frame in panel 'a', assuming a first-photon TDC-based sensor. In this dataset, the selected macropixel no longer shows a signal peak due to photon pile-up effects. To generate this depth frame, signal and ambient photon rates were estimated for each macropixel and entered into a sensor model with 10× finer TDC resolution (70 ps) and 4× narrower instrument response function (σ = 100 ps) than the present system. The same number of total laser cycles (18000) were used as in the original data, and the frame rate of single shot time stamps was taken to be 500 kFPS. Histograms were then composed from 500 exposures and peak extraction applied assuming a minimum range of 250 cm. This is to avoid the "false" peak at the start of the histogram caused by pile-up.

reduced to a level commensurate with the frame time of 1ms or shorter. A comparison with existing upscaling schemes using examples from the Middlebury dataset [28] shows generally higher accuracy than the state-of-the-art GTV [29] algorithm, together with an order-of-magnitude speed improvement, and better edge-preserving properties (see Supplementary Information).

3. DATA ANALYSIS

Data Acquisition and Depth Calculation

An Opal Kelly XEM7310 FPGA integration module is used to interface to the SPAD sensor. With the data output clock set to 100 MHz, frames are acquired at a rate of up to 1.5 kFPS, and streamed continuously over a USB3.0 link to the PC. A software interface implemented in Matlab controls the acquisition of data, and decodes the frames. Assuming a Gaussian system impulse response, depth frames are produced using an approximate maximum likelihood estimator that can be efficiently computed using a localized centre-of-mass of TCSPC histograms. In the default case, the following equation is used to estimate depth d:

$$\hat{d} = \frac{\sum_{t=\max(d_{max}-t_l,1)}^{\min(d_{max}+t_r,16)} t \max(0, h_t - b)}{\sum_{t=\max(d_{max}-t_l,1)}^{\min(d_{max}+t_r,16)} \max(0, h_t - b)}, \quad (1)$$

where h_t (t = 1...16) are the histogram bins at a given macropixel, d_{max} is the index of the bin with the maximum count and b is the median of the bins used as a measure of the ambient level. The parameters t_l, t_r are chosen such that the calculation encompasses the width of the histogram peak (typically $t_l, t_r = 2$). Due to the background compensation, and the centroid being calculated locally, the bias in the estimate is found to be minimal in simulations (see Supplementary Information). When imaging low reflectivity objects, such as the hammer head in Figure 9, in front of a background of much higher reflectivity, it is useful to extract the histogram peak closer to the camera, rather than the highest peak. We test for the existence of a second, closer peak by setting $h_t = b$ for $t = (d_{max} - t_l)...(d_{max} + t_r)$ and comparing the maximum bin count of the modified histogram with the threshold [30]:

$$h_{thresh} = b + 4\sqrt{b}, \quad (2)$$

which, under the assumption of Poisson noise on the bin counts, corresponds to a peak that is more than four standard deviations away from the baseline ambient level. If the threshold is exceeded then Equation 1 is applied to estimate the time position of this second peak, with d_{max} now corresponding to the second peak. Values of $\hat{d}$ are converted to distance via the scaling factor $c\delta/2$, where c is the speed of light and δ is the bin width (typically 700 ps). To compensate for timing skew across the sensor, which arises from clock distribution, a calibration depth frame is captured for a flat surface at a known distance, and subtracted from subsequent depth frames.

The resulting precision in the depth values is plotted in Figure 5 for increasing ambient levels. Curves are shown for target reflectivities of $<$ 10% and $\approx$ 80%, with the target at 2 m distance. The results, obtained for an exposure time of 500 μs (860 FPS), indicate sub-centimeter precision, even at the highest LED

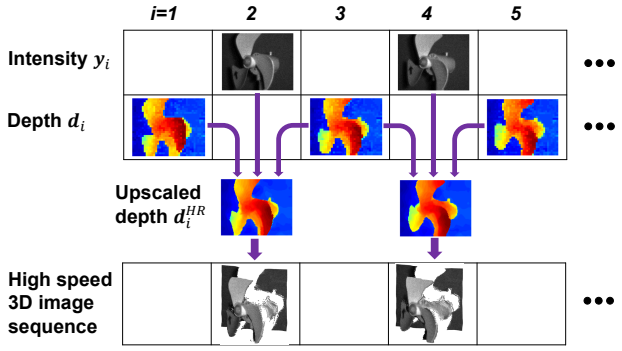

Fig. 4. 3D image construction scheme. A data sequence alternating between intensity and histogram frames is captured. The histogram frames are converted into depth by applying peak extraction to the histogram from each macropixel. By interpolating between pairs of depth frames, additional depth frames are created, aligned in time with the intensity frames. These depth frames are then upscaled in x,y, guided by the corresponding intensity data. Finally, the intensity data is overlaid onto the upscaled depth to give 3D image frames.

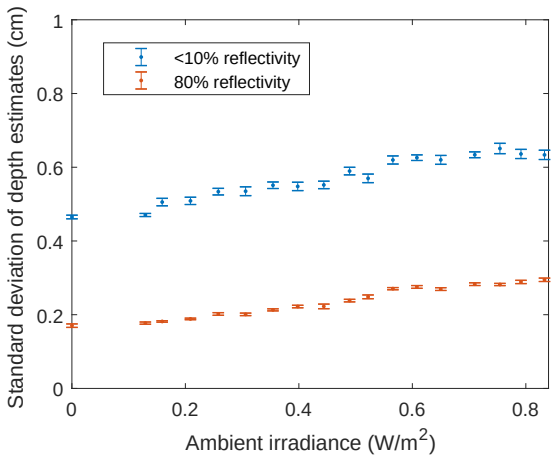

Fig. 5. Precision of depth estimates for two different target reflectivities and a range of ambient light levels (as measured at the target using a Thorlabs PM120D power meter). The distance to the flat target was 2 m. The precision is measured as the standard deviation of depth values, across 100 exposures, the median value being taken over a 10×10 macropixel region of interest in the sensor array.

setting (which was the setting used in the indoor imaging examples in this paper). The accuracy was previously characterised [24] as approximately ± 2 cm.

The limited number of bins constraints the total depth range, for example, the 700 ps bin size used here leads to a ≈ 1.7 m range. Outside of this range, aliasing or wrap-around occurs. As the present paper focuses on short-range imaging, range disambiguation is not considered in detail here. However, potential solutions include a two-step ranging approach [32] leading to a scene adaptive sensing approach [33, 34]. In the first step, the bin size is set to a suitably large size such that the entire distance range of interest is covered. Once a measure of the absolute depth has been obtained this way, we switch to a smaller bin size to track the depth with sub-bin precision at high frame rates. We note that the laser energy must spread over multiple bins for sub-bin precision to be attained. In practice, it is expected that only a small fraction of frames would need to be captured at the wide bin setting for effective range disambiguation, the impact on the effective frame rate therefore being limited. An alternative is to use solely the wide bin setting, and scale the laser pulse width (and power) accordingly. As an example, a 16 ns bin width would give a depth range of ≈ 38 m.

Depth Upscaling

The depth frames obtained as detailed above are at the macropixel resolution of camera, which at 64×32 is relatively low. To overcome this limitation, 14-bit photon counting frames are captured in alternate frames, and used to guide the upscaling of depth data to a 256×128 resolution matching that of the intensity data. This upscaling process raises several challenges due to (i) the requirement to preserve edges to avoid artificially "joining up" distinct surfaces in the scene, (ii) the possible misalignment between the depth and intensity images for rapidly varying dynamic scenes [35], and (iii) the need for fast processing approaching real-time rates. As detailed in the Supplementary Information, while there are a number of existing method which tackle these challenges separately [35–37], the aim here is to deal with all three at the same time. The proposed strategy is based on two main steps: (1) interpolate the low resolution depth maps at times corresponding to the intensity frames, (2) generate the high-resolution maps, with both steps considering the measured high resolution intensity maps, as indicated in Figure 6. Inspired by the alternating direction method of multipliers [38, 39] or regularization by denoising [40] approaches that alternate between an estimation and filtering/denoising steps, each step of our method has two sub-steps, an estimation sub-step followed by a filtering sub-step to improve performance. To ensure fast processing, the estimation is performed using analytical expressions or simple operations. Edges are preserved in the filtering step by adopting ℓ_1-norm based algorithms such as the weighted median filter [41]. Further details on the method, together with comparisons with existing upscaling approaches in simulations, can be found in the Supplementary Information.

4. RESULTS

We present illustrative results obtained with the above approach, demonstrating the high-speed capture of 3D scenes. Figure 7 shows the application of the algorithm to the outdoor juggling data in Figure 3. The results are presented in the form of intensity, depth, upscaled depth, and 3D image frames. In each case, a set of three frames are given, separated by a time interval corresponding to 30 raw (15 SPC and 15 TCSPC) camera frames. The final 3D image frames are seen to be enhanced in definition compared to Figure 3a. We note an artefact protruding from the left hand side of the person; this is due to a feature in the background matching the shade of the person's T-shirt. Figure 8 gives the results of a similar juggling sequence, but captured indoors. Comparing the upscaled depth frames (row c) with the original (row b), improved smoothness can be seen along edges in depth. This is achieved whilst preserving the edges: no obvious interpolation effects are visible between the person's hands and chest, nor between the head/shoulders and background. Furthermore, there is more detail overall on the upscaled frames, as demonstrated by the individual fingers on the hands being better defined. Figure 9 shows another set of

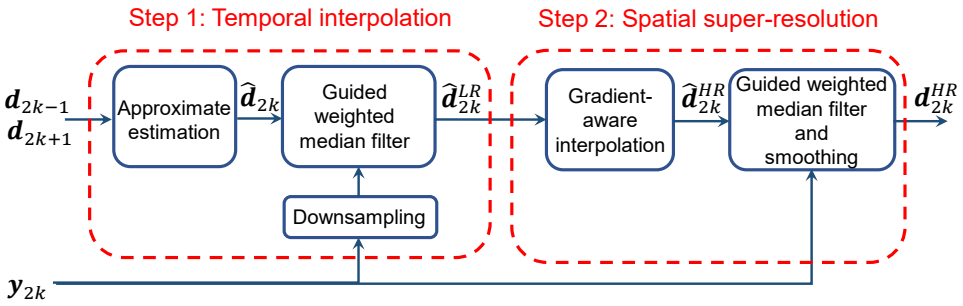

Fig. 6. Block diagram of algorithm for generating upscaled depth frames, summarising the non-iterative, two-step approach. The algorithm is also illustrated in the form of example input and output frames in Figure 4.

example frames, capturing an apple being struck with a hammer. From the intensity frames as well as the depth frames, we can readily identify individual pieces of fruit flying off with high speed following impact. The upscaled depth frames (row c) show an improvement in the definition of the edges of these pieces, at the expense of very small fragments (of a size similar to or smaller than a macropixel) being smoothed out. There is also evidence of noisy depth values (resulting from photon shot noise), as seen, for example, on the lower right corner of the first frame in row b, being removed by the upscaling process.

Videos of all the above examples can be found in the supplementary material. In addition, we show depth sequences capturing the high-speed fan at exposure levels down to 50 μs (1418 FPS), demonstrating the viability of 3D imaging at > 10 kFPS (provided a faster sensor read-out), even with the modest laser optical power currently in use.

5. DISCUSSION AND OUTLOOK

By exploiting the multi-photon timing, in-pixel histogramming functionality of a SPAD ToF image sensor, depth can be captured at frame rates above 1 kFPS. The acquisition of depth frames may also be combined, in a time-interleaved fashion, with that of higher resolution intensity frames. Whilst this halves the frame rate of native depth frames, it enables additional, upscaled depth frames to be generated, guided by and aligned with the intensity frames. We have demonstrated the practicability of the scheme in the capture of high-speed 3D sequences, even under high ambient illumination, with modest laser power requirements. This system is therefore highly relevant for applications, such as collision avoidance in robotics, where fast 3D perception that matches or exceeds human reaction times is required. To that end, we can see a number of ways that the system could be further improved:

- Whilst native depth frames can be obtained with minimal processing, upscaled depth frames currently take several 10's of milliseconds to produce. The target is to reduce this latency down to (sub-)millisecond levels.
- The current algorithm provides upscaled point depth estimates without uncertainty measures. The reformulation of the algorithm using statistical modelling tools will allow the generation of confidence maps necessary for autonomous applications.
- The limiting factor in the frame rate for the short ranges and modest field-of-view is the read out time of the sensor. Increasing the number of output lines in the sensor from

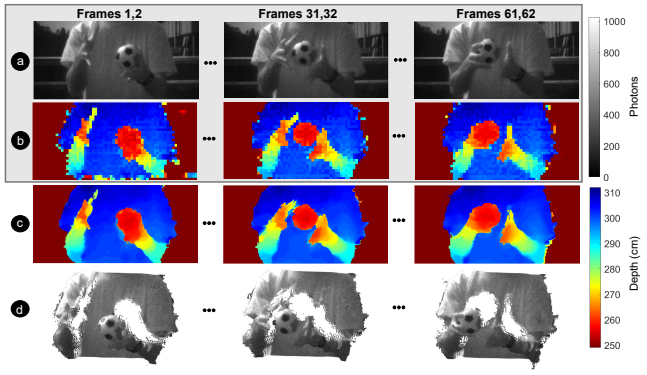

Fig. 7. Frames from a dataset of a person juggling outdoors: (a) SPC frames at 256×128 resolution (b) TCSPC frames (converted to depth) at 64×32 (c) upscaled depth (256×128) (d) upscaled depth, presented as a 3D point cloud with intensity overlaid (256×128). The exposure time was 300 μs, resulting in a frame rate of 1050 FPS. An ambient filter was used in front of the camera (Semrock LL01-671-12.5).

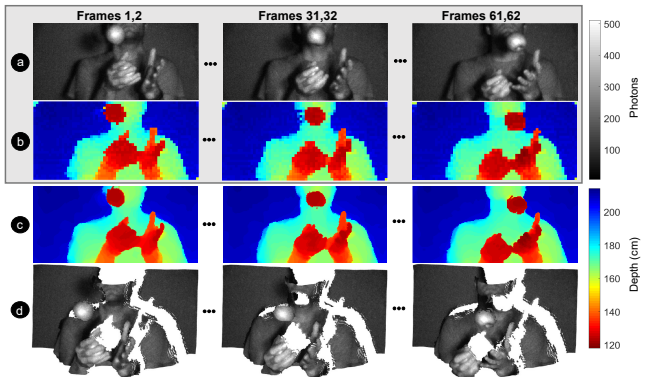

Fig. 8. Frames from a dataset of a person juggling indoors: (a) SPC frames at 256×128 resolution (b) TCSPC frames (converted to depth) at 64×32 (c) upscaled depth (256×128) (d) upscaled depth, presented as a 3D point cloud with intensity overlaid (256×128). The exposure time was 500 μs, resulting in a frame rate of 860 FPS.

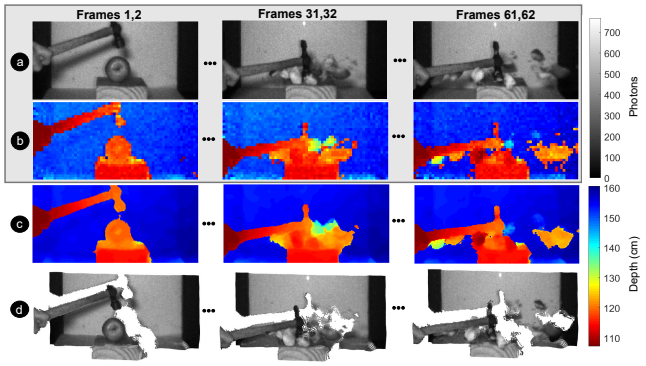

Fig. 9. Frames from a dataset of an apple being shattered by a hammer: (a) SPC frames at 256×128 resolution (b) TC-SPC frames (converted to depth) at 64×32 (c) upscaled depth (256×128) (d) upscaled depth, presented as a 3D point cloud with intensity overlaid (256×128). The exposure time was 500 μs, resulting in a frame rate of 860 FPS.

the current eight data outputs would enable even higher frame rates and/or support larger array sizes.

- We do not currently make full use of the information within the histograms. In particular, only a single peak is extracted from each histogram. We can extract either the highest peak or the peak closer to the sensor. It is anticipated that by extracting multiple peaks, as well as the widths of these peaks [42], the upscaling of depth could be further improved.

- A picosecond laser source is currently used, leading to an instrument response function that can be approximated by a Gaussian with $\sigma \approx 400$ ps. This means that the histogram bin width of $\delta = 700$ ps is within the range of $\sigma < \delta < 2\sigma$ identified in literature for optimal precision [43]. Nevertheless, it may be advantageous to switch to a nanosecond laser (typical of lidar), and adjusting the bin width accordingly, as these lasers are available in compact driver boards.

- Although the present work only considers imaging over a short range, the system is expected to be capable of high-frame rates at longer distances, provided the laser power is scaled accordingly, noting the inverse-square law governing photon returns [9].

The high-speed sensing that we present is enabled by the combination of the SPAD array sensor with high photon flux capabilities, firmware that provides high-speed hybrid imaging, and a guided upsampling approach to super-resolution. Re-configurable sensor architectures, paired with appropriate processing, could form the basis of future, "agile" 3D ToF systems, that recognise the environmental conditions, and adapt the data acquisition and illumination source accordingly to ensure optimal 3D perception.

REFERENCES

1. X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-view 3D object detection network for autonomous driving," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2017), pp. 1907–1915.
2. C. Steger, M. Ulrich, and C. Wiedemann, *Machine vision algorithms and applications* (John Wiley & Sons, 2018).
3. H. Cheng, L. Yang, and Z. Liu, "Survey on 3D hand gesture recognition," IEEE Transactions on Circuits Syst. for Video Technol. **26**, 1659–1673 (2015).
4. R. S. Sodhi, B. R. Jones, D. Forsyth, B. P. Bailey, and G. Maciocci, "Bethere: 3D mobile collaboration with spatial input," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (ACM, 2013), pp. 179–188.
5. R. Horaud, M. Hansard, G. Evangelidis, and C. Ménier, "An overview of depth cameras and range scanners based on time-of-flight technologies," Mach. Vis. Appl. **27**, 1005–1020 (2016).
6. S. Giancola, M. Valenti, and R. Sala, *A Survey on 3D Cameras: Metrological Comparison of Time-of-Flight, Structured-Light and Active Stereoscopy Technologies* (Springer, 2018).
7. S. Zhang, "High-speed 3d shape measurement with structured light methods: A review," Opt. Lasers Eng. **106**, 119–131 (2018).
8. T. G. Etoh and K. Takehara, "Needs, requirements, and new proposals for ultra-high-speed video cameras in japan," in *21st International Congress on: High-Speed Photography and Photonics*, vol. 2513 (International Society for Optics and Photonics, 1995), pp. 231–242.
9. G. M. Williams, "Optimization of eyesafe avalanche photodiode lidar for automobile safety and autonomous navigation systems," Opt. Eng. **56**, 031224 (2017).
10. A. Eisele, R. Henderson, B. Schmidtke, T. Funk, L. Grant, J. Richardson, and W. Freude, "185 mhz count rate 139 db dynamic range single-photon avalanche diode with active quenching circuit in 130 nm cmos technology," in *Proc. Int. Image Sensor Workshop*, (2011), pp. 278–280.
11. N. Krstajić, S. Poland, J. Levitt, R. Walker, A. Erdogan, S. Ameer-Beg, and R. K. Henderson, "0.5 billion events per second time correlated single photon counting using CMOS SPAD arrays," Opt. letters **40**, 4305–4308 (2015).
12. Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, and S. McLaughlin, "Lidar waveform-based analysis of depth images constructed using sparse single-photon data," IEEE Transactions on Image Process. **25**, 1935–1946 (2016).
13. J. Rapp and V. K. Goyal, "A few photons among many: Unmixing signal and noise for photon-efficient active imaging," IEEE Trans. Comput. Imaging **3**, 445–459 (2017).
14. A. M. Pawlikowska, A. Halimi, R. A. Lamb, and G. S. Buller, "Single-photon three-dimensional imaging at up to 10 kilometers range," Opt. Express **25**, 11919–11931 (2017).
15. M. Beer, J. Haase, J. Ruskowski, and R. Kokozinski, "Background light rejection in SPAD-based lidar sensors by adaptive photon coincidence detection," Sensors **18**, 4338 (2018).
16. A. Gupta, A. Ingle, A. Velten, and M. Gupta, "Photon-flooded single-photon 3D cameras," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2019), pp. 6770–6779.
17. J. Rapp, Y. Ma, R. M. Dawson, and V. K. Goyal, "Dead time compensation for high-flux ranging," IEEE Transactions on Signal Process. (2019).
18. A. Gupta, A. Ingle, and M. Gupta, "Asynchronous single-photon 3d imaging," in *Proceedings of the IEEE International Conference on Computer Vision*, (2019), pp. 7909–7918.
19. A. Maccarone, F. M. Della Rocca, A. McCarthy, R. Henderson, and G. S. Buller, "Three-dimensional imaging of stationary and moving targets in turbid underwater environments using a single-photon detector array," Opt. Express **27**, 28437–28456 (2019).
20. M. Laurenzis, "Single photon range, intensity and photon

flux imaging with kilohertz frame rate and high dynamic range," Opt. Express **27**, 38391–38403 (2019).
21. J. Tachella, Y. Altmann, N. Mellado, A. McCarthy, R. Tobin, G. S. Buller, J.-Y. Tourneret, and S. McLaughlin, "Real-time 3d reconstruction from single-photon lidar data using plug-and-play point cloud denoisers," Nat. Commun. **10**, 1–6 (2019).
22. D. Bronzi, Y. Zou, S. Bellisai, F. Villa, S. Tisa, A. Tosi, and F. Zappa, "Spadas: a high-speed 3d single-photon camera for advanced driver assistance systems," in *Smart Photonic and Optoelectronic Integrated Circuits XVII,* vol. 9366 (International Society for Optics and Photonics, 2015), p. 93660M.
23. F. Li, H. Chen, A. Pediredla, C. Yeh, K. He, A. Veeraraghavan, and O. Cossairt, "Cs-tof: High-resolution compressive time-of-flight imaging," Opt. express **25**, 31096–31110 (2017).
24. R. K. Henderson, N. Johnston, S. W. Hutchings, I. Gyongy, T. Al Abbas, N. Dutton, M. Tyler, S. Chan, and J. Leach, "5.7 a 256× 256 40nm/90nm CMOS 3D-stacked 120db dynamic-range reconfigurable time-resolved SPAD imager," in *2019 IEEE International Solid-State Circuits Conference-(ISSCC),* (IEEE, 2019), pp. 106–108.
25. S. Gnecchi, N. A. Dutton, L. Parmesan, B. R. Rae, S. Pellegrini, S. J. McLeod, L. A. Grant, and R. K. Henderson, "Digital silicon photomultipliers with or/xor pulse combining techniques," IEEE Transactions on Electron Devices **63**, 1105–1110 (2016).
26. G. Tolt, C. Grönwall, and M. Henriksson, "Peak detection approaches for time-correlated single-photon counting three-dimensional lidar systems," Opt. Eng. **57**, 031306 (2018).
27. F. Grull, M. Kirchgessner, R. Kaufmann, M. Hausmann, and U. Kebschull, "Accelerating image analysis for localization microscopy with fpgas," in *2011 21st International Conference on Field Programmable Logic and Applications,* (IEEE, 2011), pp. 1–5.
28. H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *2007 IEEE Conference on Computer Vision and Pattern Recognition,* (IEEE, 2007), pp. 1–8.
29. D. Ferstl, C. Reinbacher, R. Ranftl, M. Rüther, and H. Bischof, "Image guided depth upsampling using anisotropic total generalized variation," in *Proceedings of the IEEE International Conference on Computer Vision,* (2013), pp. 993–1000.
30. S. Gnecchi and C. Jackson, "A 1× 16 sipm array for automotive 3d imaging lidar systems," in *Proceedings of the 2017 International Image Sensor Workshop (IISW), Hiroshima, Japan,* (2017), pp. 133–136.
31. H. L. Van Trees, *Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory* (John Wiley & Sons, 2004).
32. C. Zhang, S. Lindner, I. M. Antolović, J. M. Pavia, M. Wolf, and E. Charbon, "A 30-frames/s, 252×144 SPAD flash lidar with 1728 dual-clock 48.8-ps TDCs, and pixel-wise integrated histogramming," IEEE J. Solid-State Circuits **54**, 1137–1151 (2018).
33. D. B. Phillips, M.-J. Sun, J. M. Taylor, M. P. Edgar, S. M. Barnett, G. M. Gibson, and M. J. Padgett, "Adaptive foveated single-pixel imaging with dynamic supersampling," Sci. Adv. **3** (2017).
34. A. Halimi, P. Ciuciu, A. McCarthy, S. McLaughlin, and G. S. Buller, "Fast adaptive scene sampling for single-photon 3D lidar images," in *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP),* (Guadeloupe, West Indies, 2019).
35. B. Wronski, I. Garcia-Dorado, M. Ernst, D. Kelly, M. Krainin, C.-K. Liang, M. Levoy, and P. Milanfar, "Handheld multi-frame super-resolution," ACM Trans. Graph. **38**, 28:1–28:18 (2019).
36. T. Shibata, M. Tanaka, and M. Okutomi, "Misalignment-robust joint filter for cross-modal image pairs," in *2017 IEEE International Conference on Computer Vision (ICCV),* (2017), pp. 3315–3324.
37. X. Guo, Y. Li, J. Ma, and H. Ling, "Mutually guided image filtering," IEEE Transactions on Pattern Analysis Mach. Intell. pp. 1–1 (2018).
38. S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," Foundations Trends Mach. learning **3**, 1–122 (2011).
39. M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems," IEEE Transactions on Image Process. **20**, 681–695 (2010).
40. Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (red)," SIAM J. on Imaging Sci. **10**, 1804–1844 (2017).
41. Q. Zhang, L. Xu, and J. Jia, "100+ times faster weighted median filter (wmf)," in *2014 IEEE Conference on Computer Vision and Pattern Recognition,* (2014), pp. 2830–2837.
42. J. Tachella, Y. Altmann, S. McLaughlin, and J.-Y. Tourneret, "3D reconstruction using single-photon lidar data exploiting the widths of the returns," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* (IEEE, 2019), pp. 7815–7819.
43. N. Hagen, M. Kupinski, and E. L. Dereniak, "Gaussian profile estimation in one dimension," Appl. Opt. **46**, 5374–5383 (2007).
44. I. Gyongy, S. W. Hutchings, M. Tyler, S. Chan, F. Zhu, R. K. Henderson, and J. Leach, "1kfps time-of-flight imaging with a 3d-stacked cmos spad sensor," Proc. IISW pp. 226–229 (2019).

FUNDING

Engineering and Physical Science Research Council (EP/M01326X/1, EP/S001638/1 and EP/L016753/1); UK Royal Academy of Engineering through the Research Fellowship Scheme under Grant RF/201718/17128

ACKNOWLEDGEMENTS

The authors are grateful to STMicroelectronics and the ENIACPOLIS project for chip fabrication. Portions of this work without the guided upsampling were presented at the International Image Sensor Workshop in 2019 [44].

DISCLOSURES

The authors declare no conflicts of interest.